

Course Description**CAP3794C | Data Analytics in the Cloud | 4.00 credits**

This course focuses on the role of a data analyst using cloud solutions. Students learn to utilize cloud based tools and strategies for collecting, storing, preparing, analyzing, and visualizing data for analytics.

Throughout the course, students will explore use cases from real-world applications, which will enable them to make informed decisions to effectively build data pipelines in a cloud environment. Prerequisites: CAP1788, CGS1540C, and CTS1145.

Course Competencies

Competency 1: The student will demonstrate an understanding of the cloud data collection system by:

1. Discussing how the elements of data influence decisions about the infrastructure
2. of a data pipeline
3. Illustrating how to design a data pipeline by using cloud services to meet a generalized use case
4. Identifying the risks and approaches to secure and govern data at each step and each transition of the data pipeline
5. Identifying scaling considerations and best practices for building pipelines that handle large-scale datasets
6. Discussing constraints such as scalability, cost, fault tolerance, and latency to be considered when designing and building a data collection system

Competency 2: The student will demonstrate an understanding of the cloud data storage and management system by:

1. Selecting a data storage option that matches the requirements (e.g., performance) and constraints (e.g., costs) of a given data analytics use case
2. Understanding how to select appropriate data layout, schema, structure, and format
3. Understanding the durability, reliability, and latency characteristics of the storage solution based on requirements

Competency 3: The student will demonstrate an understanding of the appropriate cloud data processing solution by:

1. Implementing the steps to process both structured, semi structured, and unstructured data formats in a data pipeline that is built using cloud technology
2. Illustrating how a specific cloud computing processing technology (e.g., Amazon EMR, Databricks) is used in big data pipelines
3. Discussing how to evaluate the different data processing cloud services for cost, scalability, and availability

Competency 4: The student will demonstrate how to implement a cloud-based analysis and visualization solution by:

1. Analyzing data by using cloud technology tools that are appropriate to a given use case
2. Implementing a data visualization solution that is aligned to an audience and data type
3. Evaluating dynamic, interactive, and static presentations of data
4. Evaluating and comparing analysis solutions based on the customer use case (streaming, interactive, collaborative, operational)

Competency 5: The student will demonstrate an understanding of efficient automation strategies by:

1. Discussing the principles of automating infrastructure deployments in cloud computing
2. Utilizing workflow orchestration tools, such as Step Functions, to construct and automate data pipelines
3. Analyzing the states and stages within workflow orchestration tools like Step Functions
4. Recognizing the significance of Continuous Integration and Continuous Deployment (CI/CD) and its application within a data pipeline context across various cloud platforms